

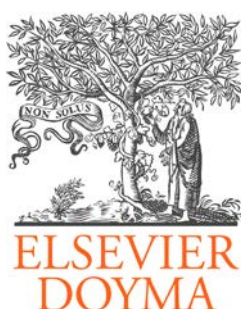


Tema 15

Revisión sistemática y meta-análisis

José Antonio González
Erik Cobo y Marta Vilaró

2014



Revisión sistemática y meta-análisis

Presentación	3
1. Introducción	4
1.1. Revisión sistemática.....	5
1.2. ¿Revisión sistemática es lo mismo que meta-análisis?	8
1.3. Fuentes de información (*)	8
2. Heterogeneidad	9
2.1. Heterogeneidad frente a variabilidad estadística	10
2.2. Heterogeneidad clínica frente a metodológica.....	11
3. Meta-análisis	13
3.1. Recogida de información. Medidas del efecto.....	13
3.2. Precisión de las medidas del efecto de los estudios	14
3.3. Estimación del efecto común	14
3.3.1. Meta-análisis con R.....	15
3.3.2. El modelo de efectos fijos.....	15
3.3.3. Medidas de heterogeneidad de estudios.....	18
3.3.4. El modelo de efectos aleatorios	20
3.3.5. Otras variables, otros análisis (*).....	25
4. Visualizando el meta-análisis.....	26
4.1. Forest Plot	26
4.2. Funnel Plot	30
4.3. L'ABBÉ	32
4.4. Covariantes **.....	33
4.5. Análisis de sensibilidad **.....	35
5. Una visión global	35
5.1. PRISMA.....	35

5.2. Crítica a las revisiones sistemáticas	37
5.3. Colaboración Cochrane	38
Soluciones ejercicios.....	42

Presentación

Este capítulo aborda las complejas circunstancias en las que se mueve una revisión sistemática. El objetivo es que el lector sea consciente sobre todo de las dificultades que implica su realización, especialmente si suponen una introducción de sesgos en la estimación del efecto de intervenciones clínicas. Los autores de revisiones sistemáticas publican sus conclusiones y, como todos los autores, pueden omitir consciente o inconscientemente limitaciones que tienen consecuencias en los resultados. Otro objetivo del capítulo es que el lector pueda interpretar correctamente la parte estadística de la revisión, conocida como meta-análisis. Con ayuda de R, contiene también las bases que permitan ejecutar análisis básicos con datos propios. Finalmente, el capítulo incluye las necesarias referencias a la declaración PRISMA, de modo que el lector la pueda utilizar como fundamento para leer críticamente una publicación basada en una revisión sistemática.

Contribuciones: José Antonio González escribió esta versión que ha sido revisada por Erik Cobo y editada por Marta Vilaró.

1. Introducción

El progreso científico, observado en detalle, no siempre es un proceso lineal. De hecho, ha evolucionado para ser así, bajo el paradigma teórico-empírico que le obliga a formularse modelos plausibles, mientras la realidad no sea tozudamente contradictoria. En ocasiones la ciencia hace un alto para mirar a su alrededor. De esto vamos a hablar en las páginas que siguen. Quien hace una revisión sistemática decide en un momento dado detener su avance individual para recapitular, examinando todo el recorrido anterior, y no solo el suyo sino —mucho más importante— el de sus compañeros de expedición.

En las ciencias de la salud, en especial, lo que revela este examen del statu quo es principalmente la cantidad de estudios realizados en torno a cierta cuestión, y la variabilidad de los resultados obtenidos por los investigadores que antecedieron a nuestro explorador. El origen de esta diversidad es múltiple: aparte de la incertidumbre propia de los participantes escogidos en cada estudio, estos fueron realizados en condiciones diferentes. El momento, el lugar o lugares, los criterios de inclusión de pacientes, las formas de administración de las intervenciones o las variables recogidas para su análisis, todo esto forma parte de la idiosincrasia de un estudio determinado.

Asumir la variabilidad presente como algo constructivo supone tratar de sintetizar la información disponible. Las preguntas fundamentales a responder son: 1) ¿Podemos encontrar evidencias de que una intervención es eficaz? Aún más: ¿Podemos aumentar la precisión disponible, y mejorar la estimación del efecto de la intervención? 2) ¿La dispersión en sí es informativa? ¿Cuáles son las fuentes de tal heterogeneidad?

Realmente, el contexto es bastante más complejo de lo que parece. La revisión sistemática no es solo un instrumento del investigador, sino que también es de interés para clínicos, gestores o administradores. Una revisión sistemática puede utilizarse para la toma de decisiones en la práctica clínica, o en la política sanitaria. Podría pensarse que es el máximo exponente de la ciencia y que, por tanto, sus conclusiones deberían ser las más firmes a las que la ciencia puede llegar. En realidad no es claramente así. Este capítulo expone las características más relevantes de las revisiones sistemáticas, sus fortalezas y sus puntos débiles, haciendo hincapié en la parte cuantitativa del mismo, que usted podrá practicar con R utilizando algunos ejemplos reales.

1.1. Revisión sistemática

Posiblemente, la primera imagen que le viene a la cabeza a un profesional clínico cuando oye la palabra “meta-análisis” o el término “revisión sistemática” es una elaborada figura como la siguiente:

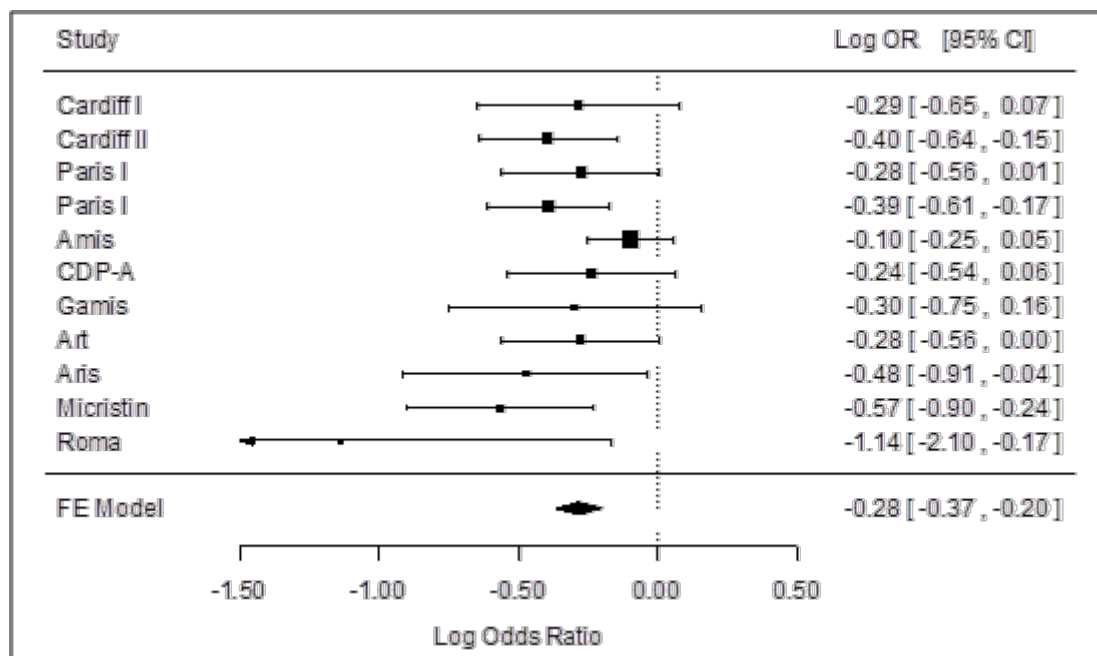


Figura 1.1. Gráfico de los resultados obtenidos en un meta-análisis

Y la mayor parte de ellos sabe que se trata de una recopilación de diferentes estudios, compartiendo un objetivo común. Por ejemplo, la Figura 1.1 busca la determinación de un efecto protector de la aspirina (o fármacos similares) contra accidentes [vasculares](#). Nuestro propósito en este capítulo es que el lector conozca los conceptos más importantes asociados al mundo de las revisiones sistemáticas y sea capaz de elaborar las suyas propias, o bien de interpretar críticamente una revisión de otros autores.

Veamos primero una definición presente en el [Diccionario de gestión sanitaria para médicos](#):



Definición

Una **revisión sistemática** es la búsqueda y evaluación crítica de todos los estudios de investigación que dan respuesta a una misma pregunta, claramente definida, que se realiza utilizando una metodología sistemática y explícita para identificar, seleccionar y evaluar críticamente las investigaciones relevantes y para recolectar y analizar los datos provenientes de los estudios incluidos en la misma.

Encontramos en la Colaboración Cochrane una interesante mención al riesgo de sesgo:

A systematic review attempts to identify, appraise and synthesize all the empirical evidence that meets pre-specified eligibility criteria to answer a given research question. Researchers conducting systematic reviews use explicit methods aimed at minimizing bias, in order to produce more reliable findings that can be used to inform decision making.

La Colaboración [Campbell](#), para educación y ciencias sociales, menciona, como requisitos para minimizar sesgos:

A systematic review must have: (1) Clear inclusion/ exclusion criteria, (2) An explicit search strategy, (3) Systematic coding and analysis of included studies; and (4) Meta-analysis (where possible)



Ejercicio 1.1

Hemos tomado algunos fragmentos tomados de diversos artículos. ¿Puede relacionar los mismos con alguno de los elementos que las definiciones previas de revisión sistemática destacan?

- a) *We conducted a systematic review* of the literature examining the association between nutrient intake, dietary components, and dietary patterns (hereafter referred to as dietary exposures) and CHD and its related clinical outcomes.
- b) *Eligible studies* were randomized controlled trials (RCTs) and cohort studies examining the effects of BMC transplantation on cardiovascular outcomes in patients with IHD.
- c) *[...] were eligible for inclusion in this study:* (i) RCTs, (ii) participants with a clinical diagnosis of AMI, (iii) the intervention consisted of any autologous BMSCs freshly isolated without restriction by dose or administration route, (iv) in the comparator arm participants did not receive BMSC (e.g. control media or plasma), and (v) co-interventions were allowed provided they were equally applied to each treatment arm. Trials were excluded on the basis of BMSCs cultured in vitro for longer than 24 h prior to infusion, as this may result in enrichment of a particular progenitor cell population.

- d) We searched MEDLINE, EMBASE, and CINAHL from inception through June 2006, and the Cochrane Database through issue 1, 2006, to identify relevant randomized controlled trials (RCTs). We also searched UMI Proquest Digital Dissertations, ISI Web of Science, and Cambridge Scientific Abstracts. We used the following search terms: pressure ulcer, pressure sore, decubitus, bedsore, prevention, prophylactic, reduction, randomized, and clinical trials.
- e) We assessed the quality [...] as A, B, or C [...]. In brief, a grade of A indicates a high-quality trial that clearly described the population, setting, interventions, and comparison groups; randomly allocated patients to alternative treatments; had low dropout rates; and reported intention-to-treat analysis of outcomes. A grade of B indicates a randomized trial with incomplete information about methods that might mask important limitations. A grade of C indicates that the trial had evident flaws, such as improper randomization, that could introduce significant bias.
- f) 19 RCTs (18 papers) in 1650 patients with IBS were identified. [...]. There were 10 RCTs involving 918 patients providing outcomes as a dichotomous variable. Probiotics were statistically significantly better than placebo (RR of IBS not improving=0.71; 95% CI 0.57 to 0.88) with a number needed to treat (NNT)=4 (95% CI 3 to 12.5). There was significant heterogeneity ($\chi^2=28.3$, $p=0.001$, $I^2=68\%$) and possible funnel plot asymmetry.

En resumen, una revisión sistemática es un enfoque alternativo a la clásica revisión narrativa, basada en el criterio del “experto”, quien utilizaba un criterio subjetivo para seleccionar los estudios recopilados.



Recuerde

La revisión sistemática dispone de un protocolo para identificar, y evaluar los estudios, analizar e interpretar los resultados, lo que facilita la transparencia y reproducibilidad de sus conclusiones.

1.2. ¿Revisión sistemática es lo mismo que meta-análisis?

Que ambos términos suelen hallarse juntos no significa que sean sinónimos (aunque algunas veces se utiliza uno en vez del otro).

El meta-análisis es solo una parte de una revisión sistemática. Meta-análisis es la técnica estadística que combina los resultados de estudios individuales para sintetizar sus resultados y dar una estimación global. El término fue introducido por Glass en 1976, en el campo de la psicología.

Revisión sistemática incluye también al proceso metodológico de búsqueda y localización protocolizada de toda la información disponible, publicada o no, con relación a la cuestión planteada sobre la que queremos investigar, así como al criterio de selección de esta información, valorando y juzgando la calidad de los trabajos encontrados.

Puede darse el caso de una revisión sistemática sin meta-análisis; sin embargo, lo contrario no tiene lógica. Podemos aplicar la técnica del meta-análisis a un conjunto arbitrario de estudios (aunque estén centrados en la misma cuestión clínica), pero si no ha habido intención de realizar una búsqueda exhaustiva, y se han elegido de forma caprichosa, la estimación resultante pierde su valor.



Recuerde

El término “revisión sistemática” hace referencia a todo el proceso, y engloba a la parte cuantitativa, conocida como “meta-análisis”.

1.3. Fuentes de información (*)

Una vez la pregunta clínica motivo de la revisión se ha formulado con rigor, y determinado la población objetivo, la intervención, los resultados medidos y las características de los estudios a analizar, debe recogerse en un protocolo los medios a utilizar para encontrar y seleccionar los estudios que entrarán en la revisión.

El principal suministrador de referencias es la base de datos bibliográficas especializadas en medicina: Medline, Embase, Cochrane Library, Pubmed, o Web of Knowledge.

Por supuesto, pueden usarse bases de datos específicas en un tema concreto. Complementariamente, puede llevarse a cabo un proceso manual, mediante las referencias presentes en los artículos hallados previamente, o consultando directamente con expertos en el área. La inclusión de la llamada “literatura gris” (trabajos no publicados o aparecidos en publicaciones de segundo orden, que los buscadores no suelen sacar a la luz) es importante, ya que contribuye a reducir el sesgo de publicación.

Tengamos en cuenta que el proceso de búsqueda idóneo se enfrenta a varios retos: maximizar la recuperación de literatura relevante y minimizar la recuperación de la irrelevante. Inevitablemente, debe hallarse un compromiso entre la cantidad y la calidad del material seleccionado en primera instancia. Construir una fórmula (utilizando lógica booleana) apropiada que emplee los términos más significativos para escoger un original es una operación delicada, de la que dependen en gran medida los resultados obtenidos.

Ejemplo 1.1: La [declaración PRISMA](#) contiene un ítem, “Búsqueda”, que pide a los autores: “Presentar la estrategia completa de búsqueda electrónica en, al menos, una base de datos, incluyendo los límites utilizados, de tal forma que pueda ser reproducible”. Además de aportar transparencia al estudio, esta propuesta persigue generar una colección de ejemplos dignos de ser imitados por otros investigadores.

Habitualmente, dos personas realizan la búsqueda de forma independiente, quienes posteriormente comparan sus resultados y se ponen de acuerdo sobre las discrepancias. Obviamente, la estrategia definitiva suele producir de forma automática un número grande (aunque manejable) de originales. La mayoría suele desecharse tras un proceso de cribado que, basado en título y resumen, puede ser muy rápido o, por el contrario, puede ser muy laborioso y exigir la lectura atenta del trabajo completo o incluso consultar a los autores. La información recogida (autores, títulos, fuente, palabras clave, resumen, etc) se guarda y mantiene con la ayuda de software especializado: por ejemplo, RevMan o Procite.



Recuerde

Una estrategia exhaustiva, protocolizada y transparente de búsqueda es un punto fuerte de las revisiones sistemáticas.

2. Heterogeneidad

Desde el momento en que se dispone de un número suficiente de estudios se pone de manifiesto que existe una enorme variabilidad en los resultados. De hecho, cualquier experto en el tema a investigar conoce de antemano que hay discrepancias notables en los precedentes, ya antes de iniciar un proceso sistemático.



Recuerde

Existe la variabilidad de resultados entre los estudios.

2.1. Heterogeneidad frente a variabilidad estadística

La variabilidad es consustancial a los estudios clínicos, por las importantes diferencias biológicas existentes entre los pacientes y participantes en general. Los ensayos clínicos controlados y aleatorizados combaten este factor, que impide distinguir el efecto de las intervenciones, mediante un número suficiente de participantes tal que probabilísticamente sea casi seguro que podremos detectar una diferencia relevante, si es que esta existe. Sin embargo, en este tipo de estudios, los participantes son seleccionados con los mismos criterios, de manera que a priori la población que entra en un brazo del estudio es igual que la de otro brazo (suponiendo que no se introducen sesgos indeseados). Idealmente, entre grupos de tratamiento no hay heterogeneidad: hablamos de variabilidad sensible al muestreo. Si por azar se hubieran elegido otros individuos los resultados habrían cambiado en el detalle, y muy probablemente los números generales no se habrían modificado sustancialmente.

Por supuesto, en muchos ensayos clínicos hay factores diferenciales entre participantes: por ejemplo, es habitual que en un estudio intervengan centros diversos, a menudo de distintos países (y continentes). Sin embargo esta diversidad es intencionada, con el fin de poder alcanzar el mayor grado de generalidad posible.

En cualquier caso, un estudio multicéntrico o internacional está coordinado y dirigido por un protocolo central. En una recopilación de estudios separados no ha existido tal coordinación, ni los objetivos estaban compartidos más que de una forma genérica.

Ejemplo 2.2: [Intervenciones para el dolor en la región lumbar](#). Este trabajo del año 2009 obtuvo material organizado según 8 técnicas distintas: *In total 83 randomized controlled trials met the inclusion criteria: exercise therapy (n = 37), back school (n = 5), TENS (n = 6), low level laser therapy (n = 3), behavioural treatment (n = 21), patient education (n = 1), traction (n = 1), and multidisciplinary treatment (n = 6)*. En vez de combinar todos los estudios de una vez, el análisis se ha efectuado por separado según cada una de las técnicas.

De todos modos, aunque mezcláramos todos los estudios hallados y obtuviéramos un resultado conjunto, ¿sería de alguna utilidad? ¿Cómo se interpretaría? Cuando tenemos intervenciones muy diferentes, combinarlas no tiene sentido desde el punto de vista clínico.

En la literatura se diferencia entre heterogeneidad clínica (por ejemplo, diferencias entre pacientes, intervenciones, respuestas, etc.), y heterogeneidad metodológica (relacionada por ejemplo con el diseño del estudio, o el origen de determinado sesgo). La heterogeneidad estadística hace referencia a la variabilidad de las estimaciones de los efectos del tratamiento estimados en los diferentes estudios, y en gran medida tiene su origen en la heterogeneidad metodológica y clínica. Por ejemplo: los ensayos clínicos que no realizan de forma adecuada la ocultación de la asignación a los grupos de tratamiento son propensos a sobreestimar el efecto de la intervención.

La heterogeneidad clínica aparece cuando esta variabilidad del efecto está ligada a un factor, característica o condición del paciente. Sin embargo, una diversidad de estas condiciones no implica a la fuerza una variación en el efecto del tratamiento. Tampoco la heterogeneidad metodológica supone necesariamente que exista un efecto del tratamiento variable.

**Recuerde**

Heterogeneidad refiere a una variabilidad del efecto observado mayor de lo esperado si solo contáramos con el azar.

2.2. Heterogeneidad clínica frente a metodológica

Una revisión puede considerar que solo tiene sentido incluir estudios de cierta calidad metodológica (puede ser más tolerante, con un análisis que estratifique por la calidad de los trabajos). De esta manera espera controlar la variabilidad que está asociada a las carencias de los estudios. Así, herramientas como la guía PRISMA son de ayuda para reconocer los puntos fuertes y débiles de los estudios candidatos. Por otro lado, existen propuestas como el [sistema GRADE](#) (del grupo Grading of Recommendations Assessment, Development and Evaluation). El sistema GRADE propone varios factores para evaluar la confianza en los resultados, de manera que según ciertos elementos esta confianza puede disminuir o aumentar.

Numerosas revisiones sistemáticas utilizan un sistema simplificado basado en el examen de ciertos criterios fundamentales, tales como si el original oculta satisfactoriamente o no la asignación del tratamiento, si el enmascaramiento es eficaz o no, etc. La [escala Jadad](#) de 1996 sistematiza el uso de criterios como estos, dando lugar a una valoración en una escala validada.

**Ejercicio 2.1**

Localice información que describa cómo se utiliza la escala Jadad.

Se le ha criticado hacer excesivo énfasis en el enmascaramiento, y ninguno sobre la ocultación de la asignación, que la Colaboración Cochrane señala como fundamental para evitar el riesgo de sesgo de selección.

Por otro lado, la posición del investigador o del clínico ante la heterogeneidad clínica es algo diferente. Cuando él o ella observa “Este trabajo presenta divergencias respecto al grueso del estado del arte”, respecto a la heterogeneidad metodológica, la pregunta que viene a su cabeza será algo como: “¿será porque contiene deficiencias metodológicas?”. Mientras que respecto a la heterogeneidad clínica la pregunta sería: “¿será porque hay en él alguna característica de los participantes que difiere del resto de los trabajos, y está asociada con la respuesta?”.

Una cosa es cuantificar un efecto estimado, incluida su variabilidad, y otra distinta sería poner el foco en los motivos por los que el efecto es variable. Por supuesto, una revisión sistemática puede adoptar cualquiera de las dos posiciones, y las dos son válidas. Pero una postura “exploratoria” tiene mayor aplicación a la toma de decisiones. Un interesante trabajo de [Gagnier et al.](#) pone de relieve que todavía hay un soporte insuficiente para la investigación en temas de heterogeneidad clínica, al contrario de la de origen metodológico. En su estudio señalan que existe poco consenso al respecto, aunque recogen una extensa colección de ideas que pueden servir para guiar un proceso más formal. Apuntamos una síntesis de esas ideas en la Tabla 2.1:

Planificación a priori
Experiencia clínica
Fundamento de las covariantes
Pensar a través de las categorías de las covariantes
Jerarquía de las covariantes
Identificación de covariantes post-hoc
Métodos estadísticos
Fuentes de datos
Interpretación

Tabla 2.1 Síntesis de ideas. [Tabla 5 de Gagnier et al.](#)

3. Meta-análisis

3.1. Recogida de información. Medidas del efecto

Habitualmente no se dispondrá de la información completa de cada participante en un estudio seleccionado para una revisión sistemática. Los autores se limitan a publicar en la revista una cantidad suficiente de estadísticos que resumen el comportamiento de las principales variables medidas. Solo recientemente están apareciendo algunas publicaciones que acompañan el texto del artículo con una base de datos.

Por supuesto, el tipo de información depende de qué tipo es la variable respuesta, y del análisis estadístico realizado. Por ejemplo, si la respuesta es dicotómica (curación/no curación, éxito/no éxito) y el estudio compara dos tratamientos A y B, la información puede venir dada de esta forma:

- núm. pacientes recibiendo tratamiento A/B (n_A/n_B)
- para cada grupo, núm. pacientes que ha tenido respuesta positiva (k_A/k_B)

o también de esta otra:

- núm. pacientes en el grupo A/B y que ha tenido respuesta positiva (a / c)
- núm. pacientes en el grupo A/B y que ha tenido respuesta negativa (b / d)

Claramente, $n_A = a + c$ y $n_B = b + d$. O en forma de cuadro:

	Resp. positiva	Resp. negativa	
Tratamiento A	$a = k_A$	b	n_A
Tratamiento B	$c = k_B$	d	n_B

Con estos cuatro números (de cualquiera de las dos formas) se dispone de la información esencial de un estudio, y se puede hallar tanto la estimación puntual del efecto del tratamiento como el cálculo de la variabilidad del estimador, que permite calcular intervalos de confianza.

Habitualmente, el parámetro de interés es el odds-ratio, o el riesgo relativo, estimado por:

$$OR = \frac{a d}{b c}$$

$$RR = \frac{k_A/n_A}{k_B/n_B}$$

Ambos casos tienen el 1 como valor neutro: no existen diferencias entre los tratamientos. Sin embargo, normalmente se toma el logaritmo natural ya que esta transformación logra una distribución más próxima a la Normal (recordemos que todo estimador presenta incertidumbre por el azar de la muestra).

3.2. Precisión de las medidas del efecto de los estudios

Las variancias respectivas son (siempre que los números sean relativamente altos):

$$V(\ln(OR)) = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$$

$$V(\ln(RR)) = \frac{\left(1 - \frac{k_A}{n_A}\right)}{k_A} + \frac{\left(1 - \frac{k_B}{n_B}\right)}{k_B}$$

Si la variable respuesta es una variable cuantitativa (por ejemplo, cambio en el índice de masa corporal de pacientes obesos), la información que se requiere de cada estudio sería la siguiente:

- núm. pacientes recibiendo tratamiento A/B (n_A/n_B)
- media muestral de la respuesta en cada grupo (m_A/m_B)
- desviación tipo muestral de la respuesta en cada grupo (s_A/s_B)

En este caso, nótese que la simple diferencia de medias (MD) no representa una medida estandarizada, de manera que pueda decirse que representa una misma magnitud en cada estudio recopilado: es habitual que los diferentes estudios definan la variable respuesta con criterios particulares (por ejemplo, la dosis del tratamiento puede cambiar de un estudio a otro). Por esta razón, en ocasiones se emplea un efecto estandarizado dado por:

$SMD = \frac{(m_A - m_B)}{s}$, donde s^2 es la variancia pooled de ambas muestras:

$$s^2 = \frac{((n_A - 1)s_A^2 + (n_B - 1)s_B^2)}{(n_A + n_B - 2)}$$

Para muestras grandes, la variancia de SMD se aproxima por $\frac{1}{n_A} + \frac{1}{n_B}$.

De esta manera, para los K estudios incluidos en la revisión sistemática, se dispone finalmente de la siguiente información:

- d_1, \dots, d_K , corresponde al estimador del efecto, estandarizado (ln OR, ln RR, SMD, etc.) o no (MD, etc.)
- v_1, \dots, v_K , corresponde a la variancia del estimador d_k .

De manera general, la variancia es inversamente proporcional al número de participantes. Los estudios más grandes suelen estar asociados a menor variabilidad atribuible al muestreo.

3.3. Estimación del efecto común

Esta propiedad es utilizada para construir un método de meta-análisis muy extendido, que nos permitirá combinar racionalmente los diferentes resultados de los estudios para obtener una estimación global



Ejercicio 3.1

En realidad, los estudios con más participantes no siempre son los que tienen más peso en el meta-análisis. Invéntese unos datos (basándose en la medida que quiera, sobre una respuesta dicotómica) en los que se observe esta situación.

3.3.1. Meta-análisis con R

Antes de abordar la cuestión anterior, introduciremos unas nociones sobre cómo usar R para realizar un meta-análisis, incluyendo el ajuste de un modelo y las representaciones gráficas más habituales. Aunque existen varios paquetes diseñados para ello, nos centramos en el package ‘metafor’, uno de los más completos de entre los disponibles. Para obtener más información acerca de otros paquetes, consulte la página que CRAN mantiene en su [web](#).

En primer lugar, debe instalar ‘metafor’. En la versión actual (1.9-1) se requiere una versión de R superior a la 3.0.0, así que actualice su programa R si todavía utiliza una versión 2.X.

En general, usted va a seguir los siguientes pasos:

1. Dispone de los datos de K estudios, a través de medidas resumen de cada grupo de intervención (por simplicidad, supondremos que cada estudio solo maneja dos grupos: A y B);
2. A continuación, se derivan las estimaciones de los efectos (d_k) con su variancias correspondientes (v_k);
3. Seguidamente, se estima el modelo que se considere oportuno;
4. Por último, se generan los gráficos que permiten examinar y validar si el modelo explica adecuadamente los datos.

3.3.2. El modelo de efectos fijos

Una posibilidad que permitiría combinar los K resultados sería promediar todos los efectos calculados, a costa de considerar todos los estudios de la misma importancia. En la estimación ponderada se sigue el siguiente procedimiento:

- Se halla un peso w_k para cada estudio, igual a $1/v_k$, $k=1\dots K$
- Se obtiene la estimación global puntual: $D = \sum w_k d_k / \sum w_k$
- Se calcula la variancia del estimador global: $V(D) = 1 / \sum w_k$
- El intervalo $D \pm 1.96\sqrt{V(D)}$ es un intervalo de confianza al 95% para el efecto de la intervención A respecto B

Nota: Todos los sumatorios van de $k=1\dots K$ si no se indica lo contrario.

Nota técnica: $\sqrt{V(D)}$ equivale también al error tipo de la estimación global.

Ejemplo 3.1: La Tabla 3.1 contiene los datos de una [revisión sistemática sobre terapias de rehabilitación en casa para pacientes que han sufrido un ictus](#). Aparecen solamente seis estudios asociados con terapias ocupacionales, simplemente a efectos ilustrativos, y para reducir la heterogeneidad del conjunto (uno de ellos posee el desafortunado nombre de “Total”, pero solo es un estudio más). Las columnas 'N' se refieren al tamaño de cada grupo de intervención, y las 'n' a los casos presentados en relación a la variable principal, deterioro en la capacidad del paciente para llevar a cabo actividades de la vida diaria, o dependencia para tales actividades, o muerte.

Estudio	Tratamiento		Control						
	<i>n</i>	<i>N</i>	<i>n</i>	<i>N</i>	<i>OR</i>	<i>IC</i>		<i>V(d)</i>	<i>w</i>
Cardiff	33	55	32	54	1.03	0.48	2.22	0.15	6.56
Nottingham 95	2	42	3	23	0.33	0.05	2.16	0.91	1.10
Nottingham 97	6	53	14	58	0.40	0.14	1.14	0.28	3.54
Nottingham 99	18	90	27	86	0.55	0.27	1.09	0.12	8.10
Glasgow	33	66	41	67	0.63	0.32	1.26	0.12	8.10
Total	106	248	56	123	0.89	0.58	1.38	0.05	20.30

Tabla 3.1 Datos de una [revisión sistemática sobre terapias de rehabilitación en casa para pacientes que han sufrido ictus](#)

La suma de los pesos (columna W) es 47.7, y la suma de los productos de W con el correspondiente logaritmo natural del odds-ratio vale -15.13. El cociente -0.317 es D, una estimación puntual del logaritmo natural del odds-ratio común, aunque es más comprensible si tomamos $\exp(-0.317) = 0.73$.



Ejercicio 3.2

Complete el meta-análisis, obteniendo la estimación del efecto (OR) por IC al 95%, e interprete el resultado.

Es muy interesante tener en cuenta que ninguno de los estudios originales había encontrado evidencias de eficacia de la terapia de rehabilitación.

Veamos qué pasa utilizando R. Primero, habilite el paquete para su uso:



Ejemplo de R

```
> library(metafor)

# Este paquete dispone de la función escalc() que permite pasar
# de los datos de los estudios (tamaños de cada grupo, medias y
# desviaciones tipo, o bien las frecuencias observadas para las
# distintas opciones de una respuesta dicotómica) a valores de  $d_k$  y
#  $v_k$ . Por ejemplo, supongamos que los datos del ejemplo anterior
# los cargamos en un data.frame llamado ocupa:

> ocupa = read.table(url("http://www-eio.upc.es/teaching
/best/datos-ejemplos/terapia-ocupa.txt"), header=TRUE)

> dat.ocu <- escalc(measure = "OR", ai = nt, nli = Nt, ci = nc,
n2i = Nc, data = ocupa, append = TRUE)

# Nótese que se ha especificado que la medida sera un Odds-
# Ratio, y que los parámetros ai, nli, ci y n2i registran las
# frecuencias observadas en las correspondientes tablas 2x2. En el
# caso del OR y del RR, se adopta la transformación logarítmica.
# Otras medidas se explican en detalle en la ayuda de escalc().
# Tras esta operación, puede tener lugar el ajuste del modelo:

> res = rma(yi, vi, data = dat.ocu, method='FE')

#  $y_i$  y  $v_i$  son las columnas que la función escalc() ha obtenido y
# dejado en el data.frame dat.ocu. El parámetro method='FE' indica
# que los datos deben ajustarse siguiendo el modelo de efectos
# fijos (Fixed Effect).
```



Ejercicio 3.3

¿Qué obtiene como salida de las funciones **escalc()** y **rma()**?

Así, ¿hemos de ver el meta-análisis como una especie de lupa, que permite amplificar efectos invisibles? Desde luego, en ausencia de heterogeneidad relevante entre los estudios, así funciona. Tal vez los estudios originales carecían de potencia suficiente para detectar un efecto clínicamente importante; en este sentido, el meta-análisis ha servido para aglutinar información dispersa y sacarla a la luz.

Así era considerado el método, cuando aún no se le llamaba por su nombre actual, y era empleado en los años 30 del siglo XX en investigaciones agrarias. En palabras de Ronald Fisher:

“When a number of quite independent tests of significance have been made, it sometimes happens that although few or none can be claimed individually significant, the aggregate gives an impression that the probabilities are, on the whole, lower than would have been obtained by chance.

It is sometimes desired, taking account only of these probabilities, not of the detailed composition of the data from which they are derived, which may be of very different kinds, to obtain a single test of the significance of the aggregate”

[Statistical Methods for Research Workers](#), pp99

Sin embargo, no se deben perder de vista las premisas de las que parte el modelo de efectos fijos, porque solo si estas se pueden considerar aceptables el resultado es válido. La principal premisa es que [el meta-análisis de efectos fijos realiza una inferencia condicional sobre los K estudios considerados](#). Para decirlo de otra forma: no es generalizable a una población diferente de estudios, es decir, a condiciones diferentes de las contempladas. Desde un punto de vista estadístico escribiríamos:

- $d_k = \vartheta_k + e_k$, es el efecto estandarizado observado del k-ésimo estudio,
- ϑ_k es el auténtico (y desconocido) efecto en tal estudio,
- e_k es una variable aleatoria, con distribución $N(0, \sigma_k)$, independiente de otros estudios, que introduce azar en las observaciones.

Por supuesto, una premisa adicional al obtener el IC es que todos los estudios comparten un mismo efecto ϑ . Pero como cualquier otra premisa, esta no es más que una afirmación que se puede poner a prueba.

3.3.3. Medidas de heterogeneidad de estudios

A continuación se presenta el estadístico Q de Cochran:

$$Q = \sum_{k=1}^K w_k (d_k - D)^2$$

En condiciones ideales, es decir, si realmente todos los estudios estiman de forma insesgada un único efecto ϑ , Q sigue una distribución de probabilidad χ^2 con K-1 grados de libertad. Esto significa que las variaciones individuales, ponderadas por la inversa de la variancia de cada estudio, se pueden atribuir simplemente al azar, mientras no sean muy grandes. Pero si algún o algunos estudios se separan del valor común de forma notable, esto se traducirá en un valor Q inaceptablemente grande. Como referencia, pensemos que el valor esperado de una distribución χ^2 con K-1 grados de libertad es precisamente K-1.

Conviene prestar atención a este estadístico: como cualquier otro que se utilice para hacer un contraste de hipótesis, no hay que llevar una conclusión estadística hasta sus últimas consecuencias. En concreto, suponiendo que la prueba se realiza con un riesgo α que nos permita decidir si

rechazamos o no la homogeneidad de los efectos, si el valor P es mayor que α ello no supone que el modelo de efectos fijos es correcto. Esto es especialmente cierto cuando el número de estudios es escaso: una de las críticas habituales a la prueba de Q es su baja potencia cuando K es pequeño, por lo que es relativamente fácil no encontrar significación estadística con Q aunque realmente exista heterogeneidad. De la misma manera, se ha observado que este estadístico posee una potencia excesiva en presencia de variabilidad negligible cuando estamos frente a un número alto de estudios. Un inconveniente adicional es que el estadístico Q puede avisarnos de que estamos enfrente de una situación con heterogeneidad, pero no de informarnos de la magnitud de la misma.

Complementariamente, suele utilizarse el siguiente estadístico:

$$I^2 = \frac{(Q - (K - 1))}{Q} \times 100\%$$

I^2 tomará el valor 0 cuando Q sea menor que $K-1$. Este estadístico, propuesto por [Higgins en 2002](#), considera que la heterogeneidad estadística es inevitable, por lo que la prueba pertinente no es ver si existe heterogeneidad, sino si existe inconsistencia. Al descartar la parte “esperada” de alguna manera estamos descontando la variabilidad atribuible al muestreo.

Con el valor de I^2 no solo tenemos una medida que se puede calificar como “baja”, “media” o “alta”, sino que es posible disponer de un intervalo de confianza para el grado de heterogeneidad entre estudios. Una aplicación directa es que si el IC contiene el 0 la homogeneidad no es descartable.

La mayor ventaja del estadístico I^2 es que es fácil de interpretar, pero comparte prácticamente los mismos problemas de potencia que Q .

**Recuerde**

Se recomienda no utilizar un procedimiento estadístico para decidir si debe usarse un modelo de efectos fijos o no: preferiblemente, la elección del modelo debería hacerse por criterios separados de los propios datos.

Ejemplo 3.2: Utilizando el Ejemplo 3.1 de las terapias ocupacionales, hallaremos el valor de Q:

<i>Estudio</i>	<i>OR</i>	<i>w</i>	<i>w(d-D)²</i>
Cardiff	1.03	6.56	0.79
Nottingham 95	0.33	1.10	0.67
Nottingham 97	0.40	3.54	1.26
Nottingham 99	0.55	8.10	0.67
Glasgow	0.63	8.10	0.15
Total	0.89	20.30	0.84
		Q =	4.396

Tabla 3.2 Datos del **Ejemplo 3.1** para ilustrar el cálculo de Q

De este análisis podemos deducir que el grado de heterogeneidad entre los seis estudios ha de ser escaso, ya que el valor de Q está por debajo de su valor esperado ($K-1=5$). Con un poco más de rigor, la probabilidad de que una variable χ^2 de 5 grados de libertad sea observada con un valor superior a 4.3960 es aproximadamente un 50%. Es decir, el nivel de dispersión entre estudios se corresponde con el esperado para estudios que tratan de estimar el mismo efecto; las diferencias se pueden atribuir al azar.

En consecuencia, por ser inferior al valor de $K-1$, el estadístico I^2 vale 0%.



Ejemplo de R

```
# Veamos cómo se puede reconstruir el valor de Q y su valor P,
# tal como salen del objeto res.

> names(res)

# Esta instrucción le indica los diversos componentes calculados
# por la función rma(): por ejemplo, res$b es la estimación
# puntual. Pruebe a hacer lo siguiente:

> W = 1/res$vi
> H = W*(res$yi - res$b)^2
> sum(H)
> 1-pchisq(sum(H), 5)
```

3.3.4. El modelo de efectos aleatorios

El modelo de efectos fijos, en situaciones reales, es prácticamente una entelequia ya que difícilmente una recopilación de originales va a hallar solamente estudios que no difieran entre sí de alguna forma relevante. Con ello no se pretende decir que los meta-análisis que aplican esta técnica carezcan de fundamento, o mucho menos que no se debería utilizar en ninguna ocasión.

Sencillamente, se trata de un sensato recordatorio de que no hay que “creerse” a pie juntillas un modelo, sino utilizarlos porque son útiles. El modelo de efectos fijos puede ser bueno por el principio de parsimonia: si explica lo que veo y lo hace de forma simple, ¿para qué buscar otro?

Para ilustrar el comentario, veamos el caso de las terapias ocupacionales. Los autores no aprecian heterogeneidad, pero de hecho había más estudios que investigaban este tipo de terapias en la revisión sistemática (la tabla de las características de los estudios incluidos hace referencia a dos más), que no se pudieron utilizar porque no aportaban la información necesaria sobre la respuesta. No sabemos si el meta-análisis ampliado a todos los estudios hubiera señalado indicios de heterogeneidad, por no ahondar en la posibilidad de que la búsqueda realizada por los autores podría haber sido más intensa.

Veamos primero qué es un modelo de efectos aleatorios desde un punto de vista más formal. Esto puede plantearse con un modelo de las observaciones recogidas. Si el modelo de efectos fijos era:

$$d_k = \vartheta_k + e_k,$$

ahora supondremos que corresponden a:

$$d_k = \mu + u_k + e_k,$$

donde μ es el efecto real promedio y u_k corresponde a la variación entre-estudios, que sigue una distribución de probabilidad $N(0, \tau^2)$.

**Recuerde**

El objetivo es estimar μ y la variancia τ^2 ; si $\tau^2=0$ entonces no hay heterogeneidad.

Un matiz fundamental en el planteamiento del modelo de efectos aleatorios es que proporciona un marco para realizar inferencia incondicional sobre una población de estudios mayor que el simple conjunto de K estudios seleccionados, que se entienden como una muestra aleatoria de aquella población. Tal población hipotética no es solo la de los estudios que se hayan realizado, sino que contempla además los que se podrían haber hecho y los que se harán en el futuro. La pregunta que se trata de responder es en realidad: ¿cuánto vale μ ? (por supuesto, la pregunta solo se puede contestar incluyendo un margen de incertidumbre, por ejemplo mediante intervalos de confianza).

Nota: Una tercera posibilidad que aparece a menudo es el modelo con efectos mixtos. Según este enfoque, el valor observado de los datos de cada estudio depende de un efecto aleatorio y de una combinación lineal determinista de una o varias covariantes x_{kj} :

$$d_k = \beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp} + e_k,$$

La variabilidad u_k se distribuye como antes, $N(0, \tau^2)$, pero ahora τ^2 se interpreta como la heterogeneidad residual, es decir, la variabilidad entre los verdaderos efectos que no ha quedado explicada previamente por las covariantes introducidas.

El primer paso en la estimación de los parámetros es estimar la variancia τ^2 . (16) cita varios métodos, entre ellos uno de los más sencillos de implementar, ya que no requiere un proceso iterativo ni cálculos complejos. El método de [DerSimonian y Laird](#) consiste en hallar:

$$t^2 = \max \left\{ 0, \frac{Q - (K - 1)}{\sum w_k - \sum w_k^2 / \sum w_k} \right\}$$

Una vez tenemos un estimador de τ^2 se realizará el proceso correspondiente al modelo elegido (de efectos aleatorios o de efectos mixtos), obteniendo el resto de estimaciones.

En el caso del modelo de efectos aleatorios, el parámetro que falta por estimar es μ , para lo que se sigue un procedimiento similar al del modelo de efectos fijos:

- Se halla un nuevo peso w_k^* para cada estudio, igual a $\frac{1}{(v_k + t^2)}$, $k=1\dots K$
- Se obtiene la estimación global puntual: $D^* = \sum w_k^* d_k / \sum w_k^*$
- Se calcula la variancia del estimador global: $V(D^*) = \frac{1}{\sum w_k^*}$
- El intervalo $D^* \pm 1.96\sqrt{V(D^*)}$ es un intervalo de confianza al 95% para el efecto estandarizado promedio de la intervención A respecto B

Ejemplo 3.3: Usaremos una parte de los [datos recogidos en un estudio donde se analiza el efecto de Tacrine](#), un medicamento para tratar la enfermedad de Alzheimer. La variable respuesta es el cambio al cabo de doce semanas (ajustando cuando la duración del tratamiento era distinta) en el componente cognitivo de la escala ADAS. Estos son los mismos datos que se utilizan en [otro estudio](#) para explorar diferentes técnicas basadas en modelos multinivel o jerárquicos. Los autores justifican que, como medida del efecto, puede utilizarse en todos los casos la diferencia de medias entre ambos grupos de tratamiento.

Los autores constatan que la diferencia final menos basal da lugar a una variable con distribución aproximadamente normal, así que la media y la desviación típica son buenos

descriptores descriptivos. Hacen notar asimismo que la corrección lineal a 12 semanas de seguimiento es la causa de las notables diferencias de magnitud y dispersión entre estudios (tal vez no ha sido una premisa acertada).

Utilizamos como medida del efecto a lo largo de los estudios la diferencia de medias, sin estandarizar (MD). La variancia de esta medida corresponde a la variancia de la diferencia de dos medias independientes:

$$V(MD) = \frac{v_T}{n_A} + \frac{v_P}{n_B}$$

donde v_T es la variancia muestral en el grupo de Tacrine, y v_P en el grupo de placebo. En primer lugar, se obtienen unos pesos como si fuésemos a ajustar un modelo de efectos fijos:

Estudio	Tacrine			Placebo			MD	v	w	$w d$	$w(d-D)^2$
	media	sd	n	media	sd	n					
Davis	2.43	9.94	103	6.47	9.23	111	-4.04				
Farlow	-0.42	6.20	75	1.71	5.99	76	-2.13				
Forette	0.70	8.82	68	3.73	11.40	53	-3.03				
Foster	-4.50	14.88	40	-3.62	14.69	41	-0.88				
Knapp	0.05	2.71	230	1.28	3.08	176	-1.23				
									13.62	-19.76	5.643

Tabla 3.3 Datos de un [estudio real](#) para ilustrar el modelo de efectos aleatorios



Ejercicio 3.4

Complete con R los valores de las columnas que han sido suprimidos de la tabla del Ejemplo 3.3. El fichero con los datos lo puede leer de este [enlace](#).

La pequeña variabilidad muestral del trabajo de Knapp, junto con su mayor tamaño de muestra, hacen que este estudio acapare más del 85% del peso total. La primera estimación, por el modelo de efectos fijos, nos da un cambio promedio global de -1.45 puntos en la escala, con un error estándar de 0.27. El valor de Q es 5.643, mayor que su promedio de 4, pero no estadísticamente significativo.



Ejercicio 3.5

¿Cree usted que hay razones que justifican el uso del modelo de efectos aleatorios?

La Tabla 3.4 muestra algunos cálculos necesarios para la nueva estimación:

<i>Trial</i>			
	w^2	w^*	$w^* d$
Davis	0.34	0.46	-1.84
Farlow	1.03	0.69	-1.47
Forette	0.08	0.25	-0.75
Foster	0.01	0.09	-0.08
Knapp	135.74	1.82	-2.24
	<i>137.19</i>	<i>3.30</i>	<i>-6.38</i>

Tabla 3.4 Cálculos necesarios para la nueva estimación

La estimación de la variancia τ^2 es $t^2 = (5.643 - 4) / (13.62 - 137.19/13.62) = 0.46$. Modificamos las variancias para cada estudio añadiendo el valor t^2 y obtenemos los nuevos pesos w^* . Obsérvese que el peso para el estudio de Knapp es ahora aproximadamente del 55%.



Ejercicio 3.6

Complete el ajuste con el modelo de efectos aleatorios: obtenga la nueva estimación por intervalo de confianza, e interprete el resultado.



Ejemplo de R.

```
# El ajuste del modelo se efectúa con la función rma() con el
# parámetro method='DL' (de DerSimonian-Laird; hay varios métodos
# alternativos, que pueden consultarse en la ayuda de la función):
> res = rma(yi, vi, data = dat.tac, method='DL')
> summary(res)

Random-Effects Model (k = 5; tau^2 estimator: DL)

    logLik  deviance      AIC      BIC
   -8.4578    5.9916   20.9157   20.1346

tau^2 (estimated amount of total heterogeneity): 0.4639 (SE =
1.1732)

tau (square root of estimated tau^2 value):      0.6811
I^2 (total heterogeneity / total variability):    29.11%
H^2 (total variability / sampling variability):   1.41

Test for Heterogeneity:
Q(df = 4) = 5.6426, p-val = 0.2275

Model Results:

estimate      se      zval      pval      ci.lb      ci.ub
```

```
-1.9317  0.5504  -3.5096  0.0004  -3.0105  -0.8529  ***
# Utilizando summary() aparecen algunas estimaciones
adicionales: de algunas no hablaremos ahora, como las medidas de
desajuste del modelo (log-likelihood, etc). El término "tau^2"
designa a la estimación  $t^2$ , y vemos que además de Q se muestra
también el indicador  $I^2$ , y el índice  $H^2$ , otra medida de
heterogeneidad. En la parte inferior, el intervalo de confianza
del efecto de tacrine respecto placebo: un descenso en la media
entre -3.01 y -0.85. Si se hubiera empleado method='FE' se
habría ajustado un modelo de efectos fijos, no se habría
considerado la heterogeneidad de los estudios, y el intervalo
sería (-1.98, -0.92).
```

3.3.5. Otras variables, otros análisis (*)

Hemos ilustrado la técnica con la ayuda del OR y del RR, cuando la respuesta es dicotómica, y con la diferencia de medias (estandarizada o no) cuando la respuesta es aproximadamente normal, o simplemente una variable continua (confiemos que el teorema central del límite hará el resto). No vamos a ser exhaustivos, pero merece la pena mencionar algunas de las otras posibilidades que pueden darse a la hora de realizar el meta-análisis.

Para respuestas dicotómicas, otro posible indicador es la diferencia de riesgos (RA):

$$RA = \left(\frac{k_A}{n_A} \right) - \left(\frac{k_B}{n_B} \right)$$

$$V(RA) = \frac{\left(1 - \frac{k_A}{n_A} \right) \frac{k_A}{n_A}}{n_A} + \frac{\left(1 - \frac{k_B}{n_B} \right) \frac{k_B}{n_B}}{n_B}$$

De estas variancias se derivarían los pesos w necesarios para llevar adelante el meta-análisis. Conviene recordar que una de las ventajas del RA es que de él se deriva inmediatamente el NNT (número necesario a tratar), uno de los indicadores que refleja mejor la eficiencia de un tratamiento.

Un indicador que recibe mucha atención, en especial en estudios observacionales (muchos de ellos dentro del ámbito de las ciencias sociales) es el coeficiente de correlación r . Podemos transformar la correlación r entre dos variables en una variable con distribución aproximadamente Normal:

$$Z = \frac{1}{2} \log \left(\frac{1+r}{1-r} \right), \text{ donde } V(Z) = \frac{1}{n-3}$$

En general, si el artículo proporciona para la comparación de grupos un estadístico y el error estándar de dicho estadístico (que presumiblemente será el estimador de cierto parámetro relevante: por ejemplo, un HRR en estudios de supervivencia), el estudio puede considerarse para el meta-análisis. Sin embargo, en muchos casos la información que se da al lector es más pobre, pues lo

único que se aporta es un estadístico y un valor de P. Es verdad que los valores de P contienen bastante información, y que se pueden agregar entre sí, por ejemplo con el método de Fisher:

$$\sum_{k=1}^K -2\log(P_k) \rightarrow \chi_K^2$$

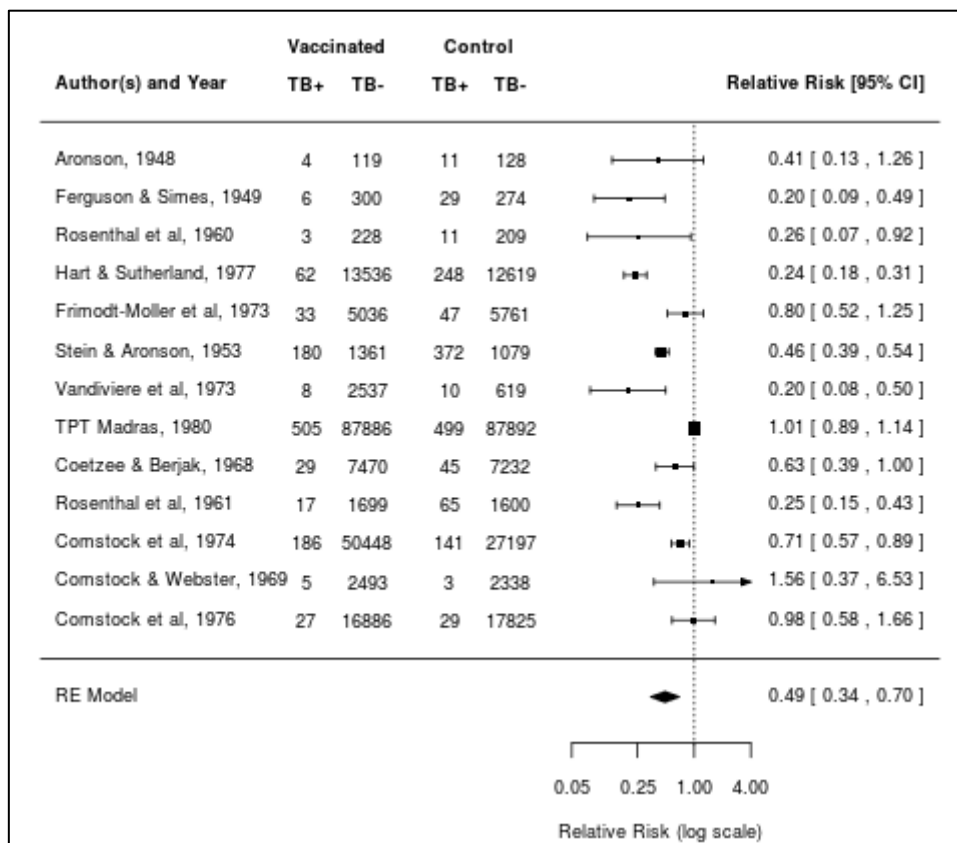
Sin embargo, este método no permite estudiar ni la magnitud ni la dirección de la asociación, y tampoco es posible analizar la heterogeneidad de los estudios.

4. Visualizando el meta-análisis

Un meta-análisis es un análisis que involucra una cantidad importante de información: K estudios, cada uno de los cuales supone un cierto número de parámetros relacionados con la respuesta de los participantes, más las estimaciones que se derivan de estos datos en relación con la medida del efecto de las intervenciones que, al fin, se agregan para obtener una estimación conjunta. Es normal que tantos números resulten oscuros para quien desee hacer un estudio y, en cambio, agradezca la existencia de algunas recomendaciones a la hora de crear representaciones gráficas inteligibles.

4.1. Forest Plot

El gráfico por antonomasia en las revisiones sistemáticas es el llamado *Forest Plot*. La Figura 4.1 muestra un ejemplo, que aparece en el [artículo de Wolfgang Viechtbauer](#). El autor no ha de preocuparse por sus derechos de imagen, ya que él mismo ha dejado en el original las instrucciones necesarias para reproducir el mismo, con la ayuda de R. En este caso, los estudios aportan información sobre eficacia de la vacuna BCG contra la tuberculosis, mostrando para grupos de vacunados y no vacunados los casos positivos y negativos.

Figura 4.1. Forest plot que aparece en el [artículo](#)

Ejemplo de R

La función **forest()** usa como único argumento el modelo ajustado por **rma()**, y dibuja un Forest Plot básico. Como suele pasar en R, es una función con otros argumentos definibles por el usuario para adaptar el gráfico a sus necesidades, junto con otras funciones que añaden otros elementos al área de dibujo, tales como textos. Este es el código utilizado para generar la Figura 4.2:

```
> forest(res, slab = dat.tac$Study, xlim = c(-25, 16), at =
c(-8, -4, 4, 8), ilab = cbind(dat.tac$m1, dat.tac$s1,
dat.tac$m2, dat.tac$s2), ilab.xpos = c(-18,-16,-13,-11), cex =
0.9)
> op <- par(cex = 0.9, font = 2)
> text(c(-18,-16,-13,-11), 7, c("m T", "s T", "m C", "s C"))
> text(c(-17,-12), 8, c("Tacrine", "Control"))
> text(-25, 7, "Study", pos = 4)
> text(16, 7, "Diff. means [95% CI]", pos = 2)
> par(op)
```

El Forest Plot es propiamente una combinación de tabla y elementos gráficos, y está constituido generalmente por:

- Una lista de los estudios implicados. A veces están ordenados de alguna forma especial: por antigüedad, por la magnitud del efecto, o por peso.
- Para cada estudio, y concretamente en cada línea, tenemos datos de identificación, datos sobre los grupos y la respuesta de los participantes, un segmento que representa el IC del efecto en ese estudio particular, y valores numéricos del efecto puntual y del mismo IC.
- El IC tiene en el centro un cuadrado lleno de área proporcional al tamaño del estudio. Así apreciamos mejor los estudios importantes, contrarrestando la impresión que producen los amplios IC de los estudios pequeños.
- Si el IC sobrepasa los límites establecidos para el eje de la medida del efecto, se dibuja una punta de flecha para indicar que el IC va más allá.
- En la parte inferior aparece un pequeño rombo lleno, que representa el IC del efecto global obtenido por el meta-análisis, junto con las estimaciones puntuales y por IC a su derecha.
- Una línea vertical marca la referencia de no-efecto; en este caso, tratándose del riesgo relativo, sobre el valor 1. Nótese que la escala es una escala logarítmica (1/4 y 4 están a la misma distancia de 1). La línea facilita la lectura de los IC, para apreciar cuáles son los estudios significativos.
- A veces también aparece información relativa a la heterogeneidad de los estudios (Q , I^2).

La gran ventaja del Forest Plot es por supuesto visual, ya que es inmediato apreciar las tendencias que puedan existir respecto a la existencia de un cierto efecto y, por otro lado, también se destaca la posible heterogeneidad entre estudios, ya que los diferentes intervalos de confianza quedarán poco o nada solapados.

El ejemplo de Tacrine aparece en la Figura 4.2. Aunque no se muestran los tamaños de los grupos, esta información se deduce del área del cuadro situado en el centro del intervalo de confianza (y queda claro que el quinto estudio es el más influyente). De todas maneras, con las herramientas de R, uno puede mostrar en el Forest Plot aquello que considere oportuno.

Otro detalle interesante es que el grado de heterogeneidad (no significativo, pero no despreciable) de este caso se manifiesta a través de una dispersión notable de las diferentes estimaciones.

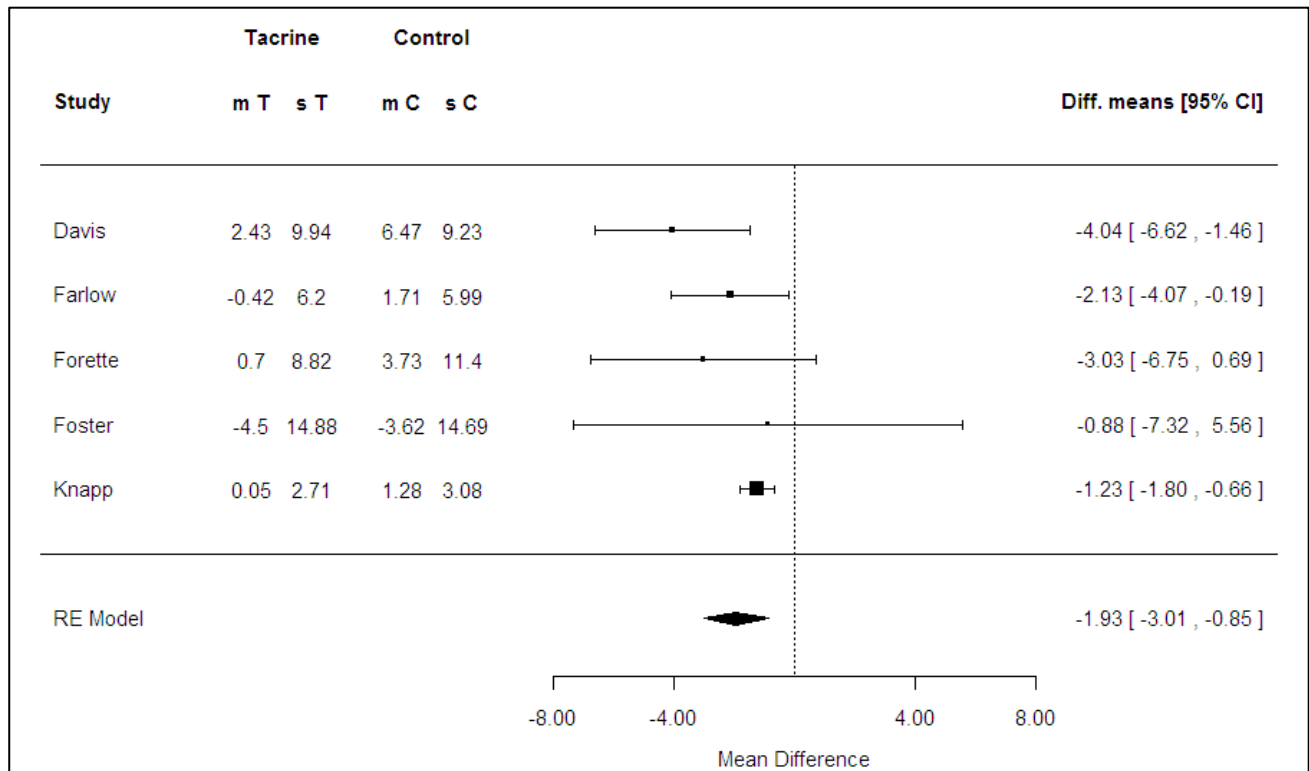


Figura 4.2. Forest plot para el ejemplo de Tacrine



Ejercicio 4.1

La Figura 4.3 muestra el Forest Plot de los datos del estudio sobre terapias ocupacionales a pacientes con ictus. ¿Qué conclusiones puede extraer de dicho gráfico? ¿Qué diría de la cuestión de la heterogeneidad de los estudios?

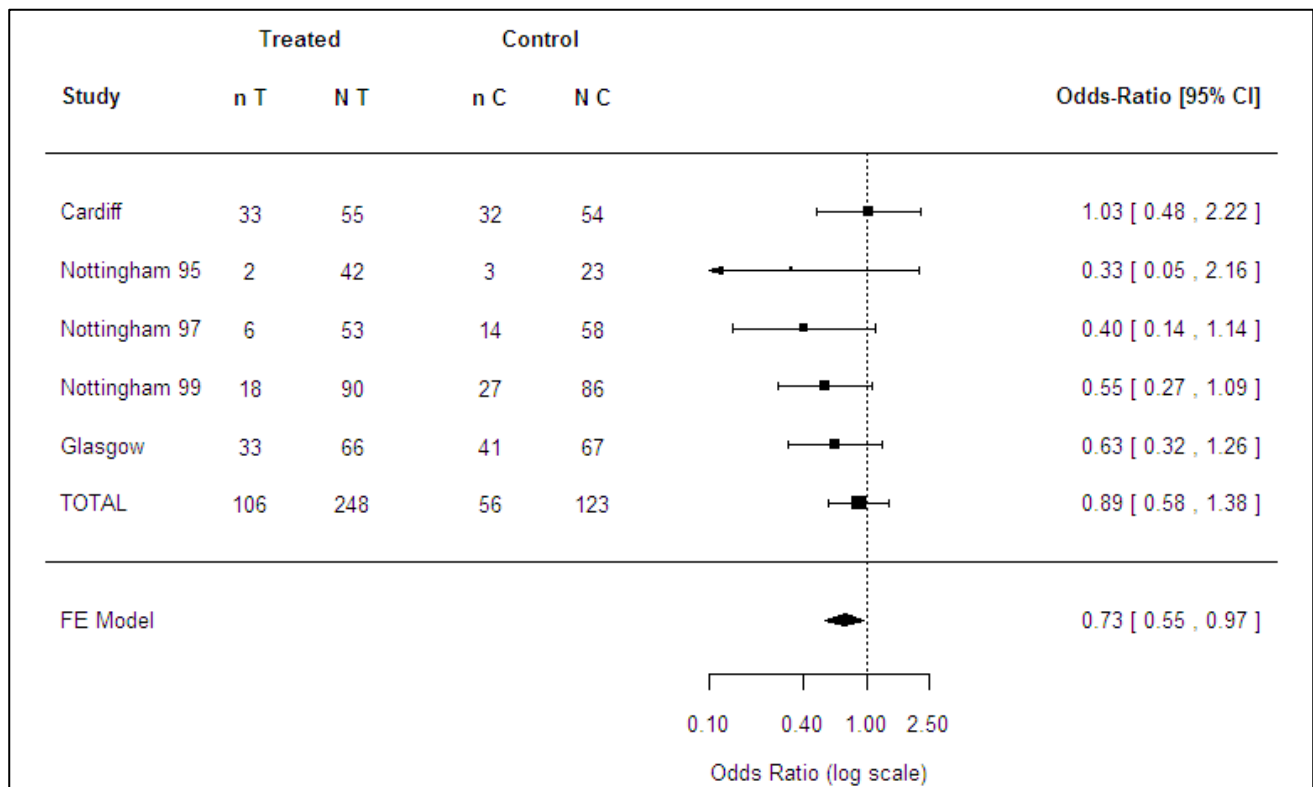


Figura 4.3. Forest plot del estudio sobre terapias ocupacionales a pacientes con ictus

4.2. Funnel Plot

El siguiente gráfico que debería ser tenido en cuenta es el *Funnel Plot*. Un Funnel Plot representa en el eje X los efectos observados en los distintos estudios, y en el eje Y alguna medida de precisión de la variable respuesta, como por ejemplo el error estándar. Si hay un número suficiente de puntos (estudios), y en ausencia de heterogeneidad y de sesgo de publicación, los puntos deberían adoptar conjuntamente el aspecto de un embudo, con la parte más ancha correspondiendo a los estudios más pequeños y menos precisos.

La Figura 4.4 corresponde a un ejemplo que acompaña al paquete metafor, acerca de los riesgos para la salud de los fumadores pasivos (trabajo de Hackshaw publicado en 1998 en *Stat Methods Med Res*). Cuando se escudriña un Funnel Plot se intenta determinar visualmente si los estudios se disponen –en sentido horizontal, ya que la dirección vertical está determinada por el número de participantes– según la lógica del azar: es decir, ninguna. En ese caso tendremos la expectativa de tener tantos puntos a la izquierda como a la derecha de la línea central, asociada al efecto común. Obviamente, no es fácil concluir que esta regla de la simetría se incumple flagrantemente. La variabilidad inherente a los estudios clínicos perturba la lectura del gráfico, en especial si no hay demasiados estudios.

**Ejemplo de R**

```
# load ETS data
> data(dat.hackshaw1998)
# fit fixed-effects model
> res <- rma(yi, vi, data=dat.hackshaw1998, measure="OR",
method="FE")
> funnel(res)
```

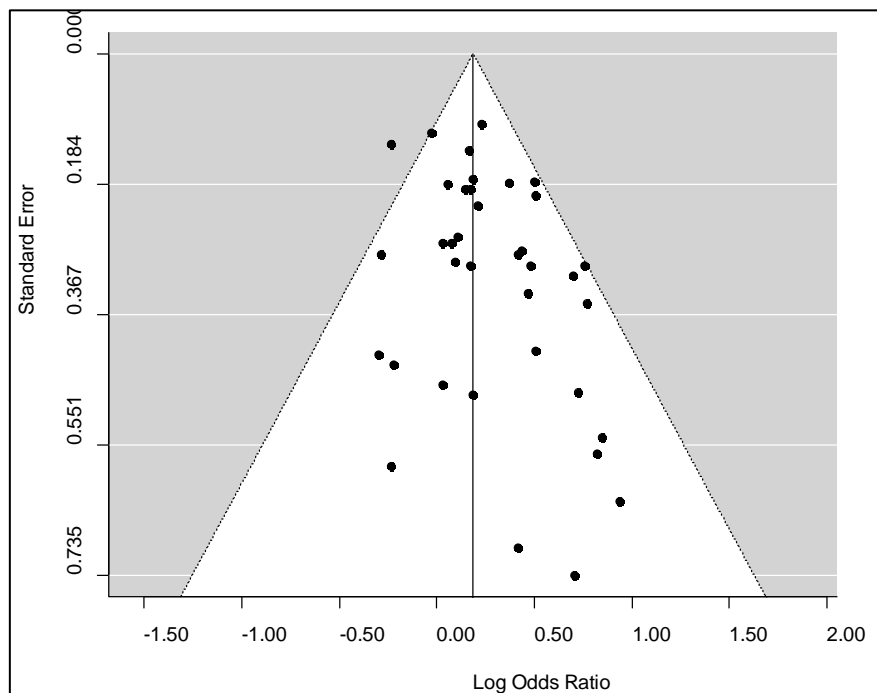


Figura 4.4. Funnel plot de ejemplo del paquete *metafor*

¿Qué puede quebrar la lógica del azar? Es decir: ¿qué podría provocar que en ciertas áreas los estudios tuvieran más tendencia a ir hacia un lado en vez de al otro? Se ha señalado que este efecto (o sesgo) tiene más propensión a producirse en la parte ancha del gráfico, entre los estudios que contaron con menos participantes, y que fueron menos costosos. Además, el lado que presenta carencias no es arbitrario: se trata del lado inesperado, el que está relacionado con efectos indeseados para el equipo investigador. Por ejemplo, para quienes (para su sorpresa) encuentran que en su propio estudio el humo de tabaco ambiental no aparece como perjudicial sino que resulta un factor protector. Un estudio sencillo y con un resultado negativo tiene más posibilidades de ser publicado solo en revistas de segundo nivel, o de no ser publicado nunca. Este es el denominado *sesgo de publicación*.



Recuerde

El sesgo de publicación es consecuencia de una asociación entre la calidad o el resultado del original y su posterior falta de visibilidad ante la comunidad científica.

4.3. L'ABBÉ

Para revisiones en las que se comparan dos intervenciones, puede emplearse el gráfico de L'Abbé. La Figura 4.5 es un ejemplo que utiliza los datos del meta-análisis de la vacuna BCG. En el gráfico de L'Abbé cada punto es un estudio y se representan los valores de la respuesta en un grupo respecto al otro. De esta forma, los puntos que se ubican sobre la diagonal representan estudios en los que el riesgo de infección no difiere entre las dos intervenciones. Cuanto más lejos esté situado un punto de esta diagonal más marcado es el efecto diferencial (cuando el eje x se utiliza para el tratamiento experimental y el eje y para la referencia, quedar por debajo de la diagonal significa que hay ventaja para el nuevo tratamiento). Los estudios se dibujan con un círculo de área proporcional al tamaño del estudio. La línea punteada paralela a la diagonal indica el efecto global estimado.

Usualmente, los estudios que se dispersan más son los menores, y los reflejados mediante los círculos más grandes se suelen encontrar más cerca de la línea punteada. En este caso vale la pena señalar el inmenso estudio MADRAS (el círculo mayor), de 1980: los autores de la [revisión sistemática](#) declaran que los números de las cohortes han sido estimados porque no fueron divulgados. Esto podría explicar en parte el nulo efecto hallado, aunque se ha visto que otros factores (por ejemplo, la latitud de la región de estudio) están asociados con la variación del efecto (ver Figura 4.6).



Ejemplo de R

```
# load BCG vaccine data
> data(dat.bcg)

# fit random-effects model
> res <- rma(ai=tpos, bi=tneg, ci=cpos, di=cneg, data=dat.bcg,
measure="RR")

> labbe(res, xlim=c(-7,-1),ylim=c(-7,-1))
> pts = c(1/1000,1/200,1/100,1/20,1/10,1/5,0.4)
> axis(3,at=log(pts),lab=pts)
> axis(4,at=log(pts),lab=pts)
```

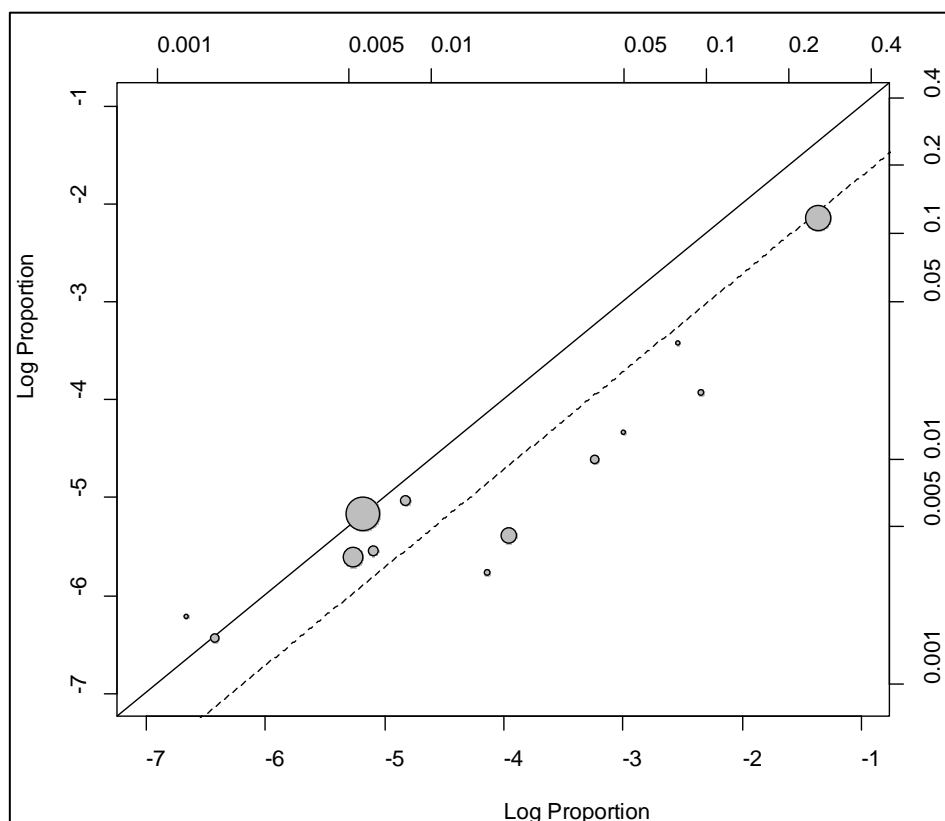


Figura 4.5. Gráfico de l'Abbé

Precisamente, el siguiente tipo de gráfico que se muestra resulta útil para explorar si determinada covariante podría influir en el resultado de los diversos estudios. En el eje de la Y se muestra el efecto para cada estudio, y en el eje de la X la variable que se utiliza como predictor. Los puntos también son de mayor o menor tamaño en función de la precisión asociada a la estimación del efecto observado. Adicionalmente, pueden incorporarse otros instrumentos visuales, tales como la recta de metarregresión o las bandas de confianza.

4.4. Covariantes **

El ejemplo que figura en la Figura 4.6 es de nuevo el de la vacuna de la tuberculosis. La covariante que se ha utilizado es la latitud absoluta de la región en donde se ha llevado a cabo el estudio. El resultado que se deduce es que la vacuna parece ser más efectiva a medida que nos separamos del ecuador. De esta manera se pueden plantear modelos de efectos mixtos, para determinar si la inclusión de una covariante (también llamadas *moderators*) puede reducir la heterogeneidad residual.



Ejemplo de R

```
> par(mar=c(5,5,1,2))
# predicciones del RR para latitudes absolutas de 0 a 60 grados
> preds <- predict(res, newmods=c(0:60), transf=exp)
# calcular tamaño de puntos según errores tipo
> wi <- 1/sqrt(dat$vi)
> size <- 0.5 + 3.0 * (wi - min(wi))/(max(wi) - min(wi))
> plot(dat$ablat, exp(dat$yi), pch=19, cex=size,
       xlab="Absolute Latitude", ylab="Relative Risk",
       las=1, bty="l", log="y")
# añadir predicciones y bandas de predicción (IC)
> lines(0:60, preds$pred)
> lines(0:60, preds$ci.lb, lty="dashed")
> lines(0:60, preds$ci.ub, lty="dashed")
# una línea en RR=1 (sin diferencias entre grupos)
> abline(h=1, lty="dotted")
```

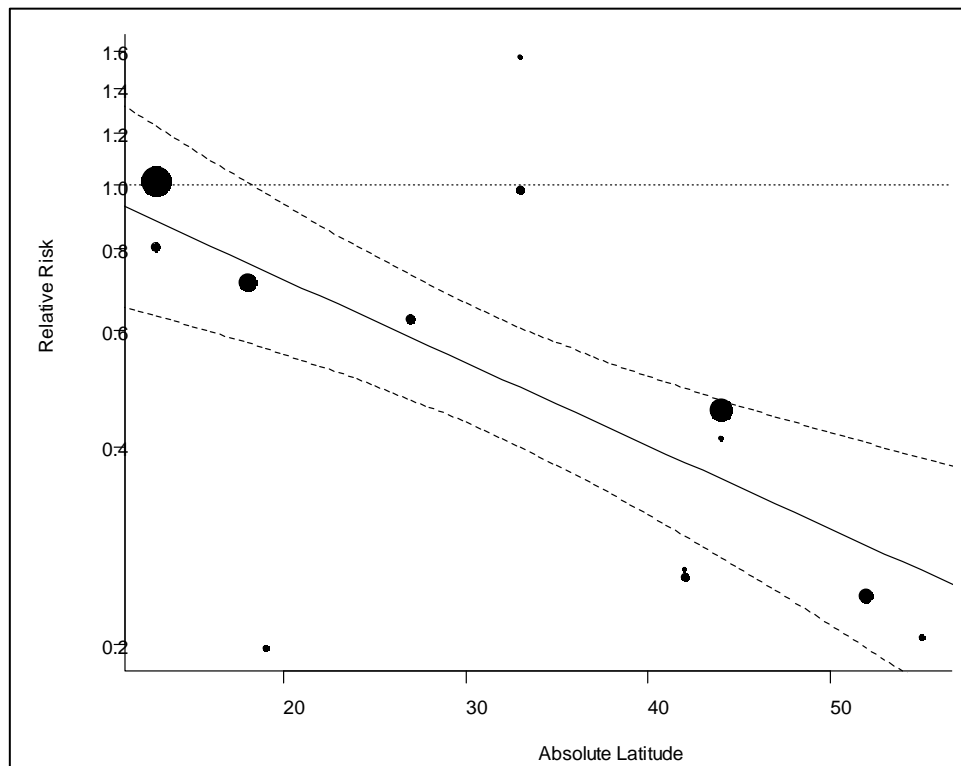


Figura 4.6. Asociación de la latitud de la región de estudio con la variación del efecto

4.5. Análisis de sensibilidad **

La exploración acerca de las fuentes de heterogeneidad entre estudios puede usar varias técnicas. Un gráfico de la librería *metafor* genera esta secuencia de 8 medidas diferentes, en donde se representan en horizontal cada uno de los K estudios. La Figura 4.7 es la aplicación de la función *influence()* para los datos de Tactine. A modo de ejemplo, citaremos el que se titula “QE.del”, que muestra el estadístico Q resultante del meta-análisis cuando vamos excluyendo de uno en uno cada estudio. Obsérvese que Q descende sensiblemente cuando se excluyen el primero (Davis) o el quinto (Knapp), puesto que son los dos que difieren más entre sí.



Ejemplo de R

```
> plot(influence(res), layout=c(2,4))
```

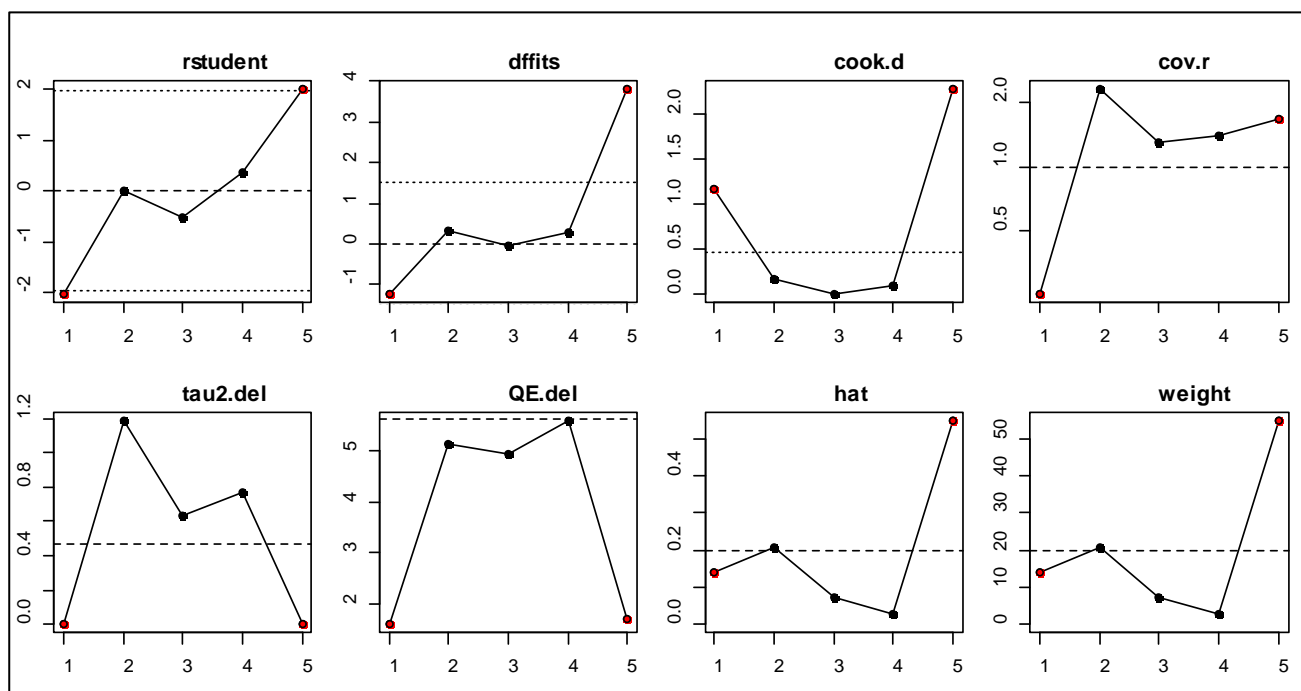


Figura 4.7. Aplicación de la función *influence()* para los datos de Tactine

5. Una visión global

5.1. PRISMA

La declaración PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) aparece en 2009, diez años después de otra declaración (QUOROM, Quality Of Reporting Of Metaanalyses), y con clara vocación de reflejar una realidad que en este periodo de tiempo había cambiado considerablemente, como se describe en la introducción de [este artículo](#).

Como otras declaraciones, PRISMA se orienta a mejorar la claridad y la transparencia de los informes publicados como trabajos científicos, cuando se ha seguido la metodología de las revisiones sistemáticas. La declaración de 2009 incluye 27 recomendaciones (frente a las 18 de QUOROM), y un diagrama en cuatro fases para describir los registros identificados, filtrados (*screening*), elegibles e incluidos. Los autores (clínicos, metodólogos, pero también editores médicos) manifiestan su voluntad de que las recomendaciones puedan ser aplicables a revisiones sistemáticas más generales que las que solo emplean ensayos clínicos aleatorizados; aunque advierten que la inclusión de estudios observacionales o con propósito ligado a pronóstico, diagnóstico o etiología (por ejemplo) pueden requerir algunas variaciones.

La lista de 27 recomendaciones está estructurada en siete apartados, de los cuales los más importantes son el de Métodos y el de Resultados. A lo largo de la lista aparece recurrentemente la mención a un acrónimo derivado de las palabras *Población*, *Intervenciones*, *Comparaciones*, *Outcomes* (respuesta) y *Study design* (diseño del estudio): PICOS. Estos términos recuerdan a los investigadores (y también a los lectores) que han de prestar especial atención a lo que se refiera a:

- Los pacientes, es decir, una definición precisa de qué tipo de participantes entran en los estudios.
- Las intervenciones a las que los participantes son expuestos, en los distintos estudios incluidos.
- El comparador, o la intervención de referencia para el nuevo tratamiento.
- La variable de respuesta utilizada en cada estudio para medir el cumplimiento del objetivo.
- El tipo de estudio, que no siempre ha de ser un ensayo aleatorizado, sino que dependiendo de la finalidad de la revisión puede considerar varios tipos de estudios observacionales.

A partir de PRISMA, el término “riesgo de sesgo” se emplea con generalidad para diferenciarse del más ambiguo “calidad” del estudio, que se venía utilizando anteriormente, dado que el término “calidad” no necesariamente implica haber puesto todos los medios al alcance para evitar la posibilidad de un sesgo. Recordemos que en un ensayo clínico aleatorizado hay unos elementos, tales como la ocultación de la asignación aleatoria o la forma de enmascarar a los agentes de la intervención, que se consideran esenciales para garantizar la validez interna del estudio: de lo contrario la posibilidad de un error sistemático destruye la credibilidad del resultado.

Una evaluación rigurosa del riesgo de sesgo también necesita una evaluación a nivel de resultados (*outcome-level*), y no solo a nivel de estudios (*study-level*). Por ejemplo, considerando la fiabilidad y validez de los datos a partir de los métodos utilizados para su medición en cada estudio particular. Por supuesto, la declaración PRISMA insiste también en la importancia de los sesgos relacionados

con la publicación, y se anima a los autores a incorporar estrategias que permitan detectar sesgos de esta especie, así como investigar su posible relación con el efecto de la intervención y la [precisión de la estimación](#).

No vamos a repetir aquí la lista de los ítems presentes en la declaración, ni a repetir los argumentos con que se justifican, ya que aparecen perfectamente expuestos en el [trabajo de Liberati, Moher y colaboradores](#).

5.2. Crítica a las revisiones sistemáticas

Meta-analyst: One who thinks that if manure is piled high enough it will smell like roses.

([Stephen Senn](#))

Los meta-análisis han recibido críticas desde bien pronto. Ha sido y es frecuente acusar al meta-análisis de combinar resultados más o menos homogéneos para producir un efecto “promedio” que al clínico le resulta de poca utilidad, ya que su interés está más cerca del paciente concreto, para el cual desea saber cuál es el [tratamiento más adecuado](#). Más allá incluso: como la selección de los trabajos relacionados con determinado objetivo ni puede ser exhaustiva ni tampoco al azar, entonces tal “promedio” está expuesto a riesgos de sesgo que pueden llevar a conclusiones dispares.

Las revisiones sistemáticas (vamos a dar por supuesto que se incluye el meta-análisis) son de hecho [estudios observacionales](#), sujetos a las limitaciones y sesgos potenciales propios de tales métodos. Aunque el material del que se nutran sean ensayos clínicos aleatorizados, no tienen por qué tener las mismas virtudes que estos: depende de qué estudios estemos poniendo juntos, de la definición de “tratamiento” en todos ellos, o los métodos para sintetizar los resultados de variables respuesta coherentes. En situación ideal, si los estudios de base son perfectamente coherentes entre sí, el meta-análisis revelará información que tal vez ninguno de los trabajos originales había demostrado. En la práctica, difícilmente se puede sortear la dificultad de tener que mezclar trabajos de diferente naturaleza, en distintos contextos, o calidad diversa.

Cuando un método es calificado como “[alquimia estadística](#)” (25) hay que suponer que más bien despierta poca confianza. Un artículo publicado en 2012 sugiere que todavía se considera las revisiones sistemáticas como investigación [secundaria o “no original”](#). Por lo menos, eso podría deducirse de su encuesta, que solo el 55% de los editores de revistas médicas de primer nivel contestaron, aunque el 71% de los que respondieron sí la calificaron como “original”, y la mayoría de esas revistas publicaron (en 2009) alguna revisión sistemática.

Pero en general se puede ver que la mayor parte de las críticas que se han levantado son [posiblemente atribuibles a las resistencias iniciales](#), y actualmente no es fácil encontrar nuevos

argumentos en su contra. El libro “[Introduction to Meta-Analysis](#)” de Borenstein et al. repasa lo fundamental de estas críticas. El lector encontrará que, en general, los defectos que se atribuyen a los meta-análisis y, por extensión, a las revisiones sistemáticas, son más bien fallos en la aplicación del método, más que problemas inherentes al método en sí. Por ejemplo:

- Un número no puede resumir un campo de investigación (sabemos que hay investigadores que se centran en la estimación del efecto de la intervención, pero no debe ignorarse la cuestión de la dispersión presente, especialmente sensible en casos heterogéneos);
- Muchos estudios no salen a la luz, y se produce un sesgo inevitable (en efecto, aunque el sesgo de publicación no es exclusivo de las revisiones sistemáticas: también se puede dar en revisiones narrativas e incluso en estudios individuales. Solamente añadir que debe propiciarse el análisis del posible impacto de este efecto);
- Se ignoran ciertos estudios importantes (suelen repetirlo quienes opinan lo contrario de lo que el meta-análisis obtiene. Hay que recordar que los criterios para decidir qué estudios son considerados y cuáles no deben establecerse previamente, y son transparentes);
- Un meta-análisis puede no coincidir con un gran ensayo clínico (se dice que esto ocurre 1 de cada 3 veces. También podríamos preguntarnos cuántas veces un ensayo clínico discrepa del siguiente ensayo que aborda la misma cuestión clínica. Pero la argumentación falla principalmente en torno a lo que se considera “coincidir”, especialmente si se refiere a conseguir un valor p menor que 0.05. Si hay conflicto, solo puede resolverse con un examen atento sobre las diferencias que se han presentado).



Recuerde

Algunos de los problemas que se han apuntado en el uso de revisiones sistemáticas tienen una base real; otros provienen de cierta dificultad metodológica y de la ligereza de uso de algunos investigadores que aplican técnicas sin conocerlas suficientemente, algo que puede y debe combatirse con más didáctica.

5.3. Colaboración Cochrane

No se puede tratar el tema de las revisiones sistemáticas sin mencionar la existencia de la Colaboración Cochrane (CC). Existe una página en internet que alberga el [Centro Cochrane Iberoamericano](#) (físicamente ubicada en el Hospital de la Santa Creu i Sant Pau de Barcelona), con amplia información sobre la Colaboración Cochrane y, en concreto, sobre la Red Iberoamericana, unos 20 países en total. La declaración PRISMA ha adoptado las definiciones CC y,

recíprocamente, muchos cambios incorporados en PRISMA también se reflejan en el Manual del revisor Cochrane.

Como menciona en su introducción, “La Colaboración Cochrane es una organización internacional, independiente y sin ánimo de lucro, establecida en el Reino Unido.”, orientada a promover y producir la información de alta calidad dedicada a orientar la toma de decisiones en el ámbito de la salud, una toma de decisiones obviamente basada en la evidencia. Dicha información debe ser rigurosa, periódicamente actualizada y fácilmente disponible. El instrumento principal para estudiar los efectos de las intervenciones sanitarias son las revisiones sistemáticas de ensayos clínicos “y otros estudios confiables”.

El acceso a las revisiones sistemáticas que mantiene la CC es abierto. Puede accederse desde <http://www.bibliotecacochrane.com>, o desde <http://summaries.cochrane.org> para acceder a los resúmenes gratuitos de un número mayor de estudios. El acceso al estudio completo por esta vía puede estar limitado a suscriptores.

La política de la CC propone a los autores un modelo muy concreto recogido en un documento oficial que se revisa frecuentemente: [*Cochrane Handbook for Systematic Reviews of Interventions*](#). En él se establece, por ejemplo, que “*Cochrane reviews, and protocols for reviews, are prepared in the Cochrane Collaboration’s Review Manager (RevMan) software and have a uniform format*”.

Acerca de RevMan, se trata de un programa de libre distribución, pensado para que los autores de revisiones según el estilo Cochrane puedan preparar más sencillamente el texto, las tablas, los gráficos y el propio meta-análisis. Para usuarios registrados, se proporciona también ayuda en su uso.

RevMan puede funcionar aisladamente, o puede interaccionar con Archie (curiosidad: Archie es el nombre de pila del médico e investigador Archie Cochrane, que dio un impulso definitivo a la epidemiología como ciencia). Archie es el repositorio en internet de la CC, que contiene los datos de las personas involucradas con la Colaboración, y la documentación y revisiones producidas. Junto con RevMan, componen el IMS (Information Management System, sistema de gestión de la información) de Cochrane, con el fin de dar soporte a los colaboradores en la tarea de producir revisiones sistemáticas de calidad que puedan ser diseminadas en la Cochrane Library o en otra parte.

Un punto muy importante, y que probablemente distingue a estas revisiones de otras destinadas a ser publicadas en una revista científica, es el interés de la CC por el mantenimiento de las revisiones. Dentro del Manual Cochrane existe un capítulo dedicado a este aspecto, y dice “La

política de la Colaboración Cochrane es que las revisiones Cochrane de Intervenciones se deben actualizar a los dos años, o incluir un comentario que explique por qué no se ha hecho;”. Una actualización puede consistir en la inclusión de nuevos estudios, o cualquier otro cambio (llamados entonces *enmiendas*), como por ejemplo (1) un cambio en la metodología; (2) la corrección de un error de ortografía; (3) volver a escribir la sección de Antecedentes; (4) la inclusión completa de un estudio que figuraba como “pendiente de clasificación”; o (5) el cambio de conclusiones al descubrir un error importante de codificación de los datos. Por supuesto, al ser posible que coexistan varias versiones de una revisión, debe procederse con cuidado para citar la versión correcta a la que se haga referencia.

Las revisiones Cochrane tienen un formato muy estructurado, tanto para protocolos como para las revisiones propiamente dichas. A continuación aparece la lista de secciones que definen una revisión Cochrane (hemos omitido los elementos no obligatorios, la lista completa aparece en el Manual):

Título*

Información de la revisión:

Autores*

Persona de contacto*

Fechas*

Resumen:

Antecedentes*

Objetivos*

Estrategia de búsqueda*

Extracción y análisis de los datos*

Resultados*

Conclusiones de los autores*

Resumen en términos sencillos:

Título en términos sencillos*

Texto resumido*

La revisión:

Antecedentes*

Objetivos*

Métodos:

Criterios de selección de estudios para esta revisión:

Tipos de estudios*

Tipos de participantes*

Tipos de intervenciones*

Tipos de medidas de desenlace*

Métodos de búsqueda para la identificación de estudios*

Extracción y análisis de los datos*

Resultados:

Descripción de los estudios*

Riesgo de sesgo de los estudios incluidos*

Efecto de las intervenciones*

Discusión*

Conclusiones de los autores:
Implicaciones para la práctica*
Implicaciones para la investigación*
Sobre el artículo:
Declaraciones de interés *

Tabla 5.1 Recuadro 2.2.b del [manual Cochrane de revisiones sistemáticas](#)

Soluciones ejercicios

1.1. Hemos tomado algunos fragmentos tomados de diversos artículos. ¿Puede relacionar los mismos con alguno de los elementos que todas las definiciones de revisión sistemática destacan?

- La cuestión clínica
- Criterios de elegibilidad
- Otra vez criterios de elegibilidad
- Estrategia de búsqueda
- Evaluación de calidad
- Combinación de resultados (o meta-análisis); suele explicarse extensamente en el cuerpo del artículo, pero el abstract suele proporcionar un resumen compacto.

2.1. Según Wikipedia, un original se valora según tres preguntas:

- ¿Se describe el estudio como aleatorizado?
- ¿Se describe el estudio como doble-ciego?
- ¿Hay una descripción sobre las retiradas y las pérdidas de participantes?

Por supuesto, cada punto debe estar correctamente justificado en el artículo. Según ello, se consideran premios o penalizaciones que dan lugar a una valoración entre 0 y 5.

3.1.

Estudio 1				Estudio 2			
	posit	negat			posit	negat	
Trat.	10	200	210	Trat.	40	80	120
Control	5	90	95	Control	30	75	105

El primer estudio incluye 305 participantes, y el segundo 225. Si la medida es el $\ln(\text{OR})$, hallamos las variancias de sendos estimadores, y obtenemos para la tabla del estudio 1 $v_1 = 0.316$ ($w_1 = 3.16$), y para el estudio 2 $v_2 = 0.0842$ ($w_2 = 11.88$). En este caso observamos que la prevalencia en el primer estudio es mucho menor que en el segundo (independientemente del brazo tratado), algo que clínicamente puede significar que ambos estudios son consustancialmente heterogéneos, no porque los OR sean muy diferentes (0.9 y 1.25 respectivamente), sino porque hay motivos para creer que la pregunta o las condiciones clínicas son distintas.

3.2. La estimación por IC al 95% se obtiene según:

$$\left[e^{\left(-0.317 - \frac{1.96}{\sqrt{47.7}}\right)}, e^{\left(-0.317 + \frac{1.96}{\sqrt{47.7}}\right)} \right] = [0.55, 0.97]$$

Observamos que, de acuerdo con esta estimación, el efecto de la intervención es estadísticamente significativo: la terapia ocupacional reduce el número de eventos respecto a la terapia control con un odds-ratio de entre 0.55 a 0.97.

3.3.

```
> dat.ocu
      Study  nt  Nt nc  Nc      yi      vi
1   Cardiff  33  55 32  54  0.0308 0.1525
2 Nottingham-95  2  42  3  23 -1.0986 0.9083
3 Nottingham-97  6  53 14  58 -0.9133 0.2821
4 Nottingham-99 18  90 27  86 -0.6046 0.1234
5   Glasgow   33  66 41  67 -0.4555 0.1235
6      Total 106 248 56 123 -0.1130 0.0493
```

Lo más importante que se debe resaltar son las dos últimas columnas: y_i representa el $\ln(\text{OR})$ para cada estudio, lo que hemos denominado el efecto d_k ; v_i es la variancia de dicha estimación. Por ejemplo, el estudio “Total” puede describirse con esta tabla:

	positivos	negativos	
Tratados	106	142	248
Controles	56	67	123

El OR vale $106 \cdot 67 / (56 \cdot 142) = 0.8931$ (cuyo logaritmo es -0.113047), y la variancia resultante se deriva de:

$$1/106 + 1/67 + 1/56 + 1/142 = 0.049259$$

```
> res
Fixed-Effects Model (k = 6)
Test for Heterogeneity:
Q(df = 5) = 4.3960, p-val = 0.4939
Model Results:
estimate      se      zval      pval      ci.lb      ci.ub
-0.3171    0.1448   -2.1902    0.0285   -0.6009   -0.0333      *
```

Podemos identificar los elementos de la última línea: el valor de la estimación puntual D (-0.3171): obsérvese que corresponde a una estimación del logaritmo del Odds-ratio; “se” equivale a $V(D)^{1/2}$ (0.1448); “zval” representa la estimación reducida, es decir, dividiendo por el error tipo ($-0.3171/0.1448$); “pval” es el valor P de la prueba que contrasta si el parámetro efecto podría ser 0; y a continuación tenemos los límites del intervalo de confianza al 95% que, si deshacemos la transformación, se convierten en $\exp(-0.6)=0.5483$ y $\exp(-0.03)=0.9672$.

3.4.

```
> tacr = read.table(url("http://www-eio.upc.es/teaching/best/datos-
ejemplos/tacrine.txt"), col.names=c('Study', 'm1', 's1', 'n1', 'm2', 's2', 'n2'),
header=FALSE)
> dat.tac <- escalc(measure = "MD", m1i = m1, sd1i = s1, n1i = n1, m2i = m2,
sd2i = s2, n2i = n2, data = tacr, append = TRUE)
> W = 1/dat.tac$vi
> w.d = W*dat.tac$yi
> D = sum(w.d)/sum(W)
> H = W*(dat.tac$yi-D)^2
> cbind(dat.tac$vi, W, w.d, H)
```

	W	w.d	H
[1,]	1.72676186	0.57911865	-2.33963936 3.88171800
[2,]	0.98463991	1.01559970	-2.16322736 0.46820156
[3,]	3.59608135	0.27808047	-0.84258383 0.69330155
[4,]	10.79867951	0.09260392	-0.08149145 0.03019509
[5,]	0.08583087	11.65081986	-14.33050843 0.56915548

Recuerde que R ha calculado las variancias v_i , de acuerdo con la medida MD: como suma de variancias de las respectivas medias: por ejemplo, para el primer estudio:

$$\frac{9.94^2}{103} + \frac{9.23^2}{111} = 1.726762$$

3.5. Hay buenas razones para adoptar el modelo de efectos aleatorios. Primero, la mención anterior de que se ha introducido un factor distorsionante en los estudios adaptando las diferentes variables respuesta a un horizonte de tiempo fijo, como si la progresión fuese lineal. Segundo, no parece prudente que un solo estudio, aunque tenga el 42% de los participantes, tenga un peso del 85%, y esta desproporción afecta necesariamente a la estimación de Q. Un modelo de efectos aleatorios redistribuirá los pesos para cada estudio de manera más ecuánime.

3.6. La segunda estimación puntual del efecto de Tacrine es $-6.376/3.30 = -1.93$, con un error estándar de 0.55 (por tanto, continua teniendo un efecto significativamente diferente de 0). Un intervalo de confianza al 95% para medir el posible efecto promedio en el descenso sobre el valor basal de la escala empleada se sitúa entre 0.85 y 3.01 puntos.

4.1. Se ve claramente que, aunque ninguno de los estudios individuales muestra significación estadística (que se demuestra porque los IC cortan la línea vertical), la compilación final obtenida con el modelo de efectos fijos (FE Model) no incluye el valor de 1. También es visible que las diferentes estimaciones no difieren mucho entre sí, lo que justifica que no haya evidencias de heterogeneidad. Si bien otra opinión podría objetar que los IC en realidad se solapan porque hay mucha variabilidad intra-estudio y, por ello, los intervalos son amplios.